IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 4, December 2023, pp. 1534~1542
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i4.pp1534-1542

Toddler monitoring system in vehicle using single shot detector-mobilenet and single shot detector-inception on Jetson Nano

Kok Jia Quan[1], Zamani Md Sani[1], Tarmizi Bin Ahmad Izzuddin[1], Azizul Azizan[2], Hadhrami Abd Ghani[3]

[1]Department of Mechatronics Engineering, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Department of Advanced Informatics, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[3]Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Kota Bharu, Malaysia

Article Info

Article history:

Received Aug 1, 2022
Revised Feb 14, 2023
Accepted Mar 10, 2023

Keywords:

Artificial intelligence
Human detection
Neural network
Tensorflow
Vision-based system

ABSTRACT

Road vehicles are today's primary form of transportation; the safety of children passengers must take precedence. Numerous reports of toddler death in road vehicles, include heatstroke and accidents caused by negligent parents. In this research, we report a system developed to monitor and detect a toddler's presence in a vehicle and to classify the toddler's seatbelt status. The objective of the toddler monitoring system is to monitor the child's conditions to ensure the toddler's safety. The device senses the toddler's seatbelt status and warns the driver if the child is left in the car after the vehicle is powered off. The vision-based monitoring system employs deep learning algorithms to recognize infants and seatbelts, in the interior vehicle environment. Due to its superior performance, the Nvidia Jetson Nano was selected as the computational unit. Deep learning algorithms such as faster region-based convolutional neural network (R-CNN), single shot detector (SSD)-MobileNet, and single shot detector (SSD)-Inception was utilized and compared for detection and classification. From the results, the object detection algorithms using Jetson Nano achieved 80 FPS, with up to 82.98% accuracy, making it feasible for online and real-time in-vehicle monitoring with low power requirements.

Corresponding Author:

Zamani Md Sani
Department of Mechatronics Engineering, Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, Durian Tunggal, 76100, Melaka
Email: zamanisani@utem.edu.my

1. INTRODUCTION

In this age, automobiles were the primary mode of human mobility. People are worried on safety aspects in automotive specs, such as the car's construction, car seat and airbag, in this instance. However, the most crucial issue in safety is the fact that individuals or drivers tend to disregard the behaviours of other passengers. Human activity recognition (HAR) is a field of study that seeks to identify a person's actions based on sensor or camera observation [1]–[4].

Due to the development of autonomous vehicles recently, a lot of research on the surrounding monitoring of vehicles has been done [5], [6]. However, the tracking of a passenger is still less to be concerned about. An accident is sometimes unpredictable and unpreventable, and not only because of the carelessness or drowsiness of the driver. What people must do in an accident is know how to survive in the accident. According to the child accident prevention trust organization, twelve children under ten are killed or injured as passengers in cars every day. An inside car safety monitoring system for toddlers is a system that recognizes the seatbelt

condition of a toddler to alert parents. The focus of this research is to develop a relatively new system using artificial intelligence which can recognize the seatbelt condition of toddlers in the backseat.

2. RELATED WORK

As a means of providing a concise introduction to object detection, this section provides an overview of relevant research on the topic of seatbelt detection difficulties. Object detection [7] is widely regarded as one of the most crucial challenges facing the area of computer vision. Every day, there is a further expansion of the scope of the object detection difficulties. In order to solve these issues, the research and development groups frequently make use of cutting-edge methods such as machine learning [8], [9].

As a subfield of artificial intelligence [10], [11], computer vision [12], [13] is described as the process of teaching computers to comprehend the visual environment. It is able to identify all of the items or persons in a picture by utilizing a mix of information and can do so with a level of success that is reasonable [14]. Through the utilization of digital image capture via cameras and learning models, the computer is able to effectively detect and discriminate between items. Computer vision has been able to emulate humans in several tasks linked to recognizing and labelling things [15], thanks to advancements in deep learning [16], [17] and neural networks. This was previously impossible. Pattern recognition is the name of the game here, and this is carried out by teaching a computer how to recognize different kinds of visual input. The autonomous vehicle, often known as a self-driving automobile, is an example of one of the more well-known applications of computer vision [18]. Computer vision is sometimes referred to as "perception" in the area of autonomous cars since cameras are one of the primary instruments that a vehicle uses to perceive its environment.

The first simulation of perceptron a was carried out by Frank-Rosenblatt on an IBM 704 computer. This ultimately resulted in the building of an electronic machine [19]. An area of artificial intelligence known as machine learning enables computers to learn from previous data or experiences without being explicitly programmed [20]. Developing computer systems that have access to data and can learn from the data they have acquired themselves is the primary emphasis of machine learning. Identifying a pattern in a big dataset is one of many applications for machine learning, which may be used in a variety of industries. The generation of example data is the initial stage of machine learning, which involves the collection and preparation of data. After then, the data that has been prepared will be input into the machine in order to train it. Following the completion of the training procedure, a model will be implemented. Creating additional example data could make the model better in the long run. The process of machine learning is illustrated in Figure 1.

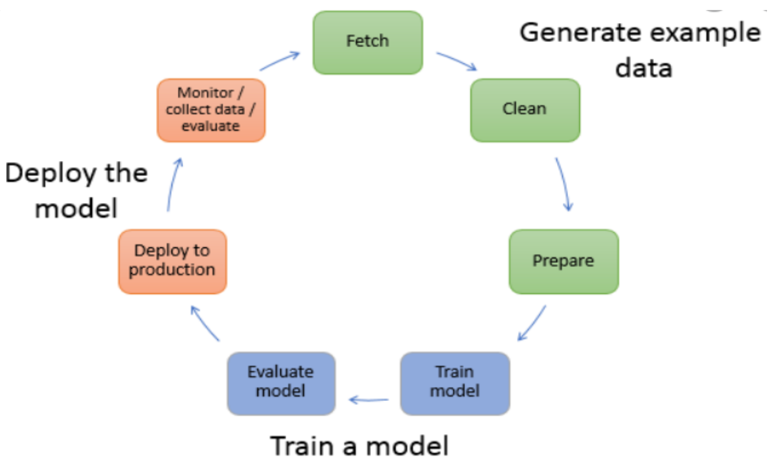


Figure 1. Flow of machine learning

Neural Network is popular for the human detection system. Yan *et al.* used the region based convolutional neural network (R-CNN) to recognize the driver's behaviour based on convolutionan neural network (CNN) whereas Nikouei *et al.* and Bao *et al.* using lightweight convolutional neural network [21]–[23] for realtime human detection and gender edge estimation. Murthy *et al.*, Yan *et al.* and Jose *et al.* [24]–[26], used convolution neural network for human pose estimation, drowsiness detection system and face recognition The hardware used are from the computer using the NVIDIA graphics processing units (GPU) Rasberry Pie and Jetson Nano for the image processig and classification. Nevertheless, most of the past research did emphasized on the detection process at the front or driver seat. There is less concern about the passenger,

especially the toddler. Furthermore, the safety of the passenger is not stated in the journals. From these gaps, a toddler monitoring system in the vehicle by using artificial intelligence will be developed.

3. METHODHOLOGY

In this section, SSD-Inception and SSD-Mobilenet are used as the networks for the toddler monitoring system. Both networks are trained to detect the toddler's situation in the backseat. More specifically, the main objective is to split up three cases: (1) detect the presence of a toddler, (2) classify the safety condition by detecting the seatbelt, and (3) compare the performance of the networks. The flow for the monitoring system is shown as in Figure 2.

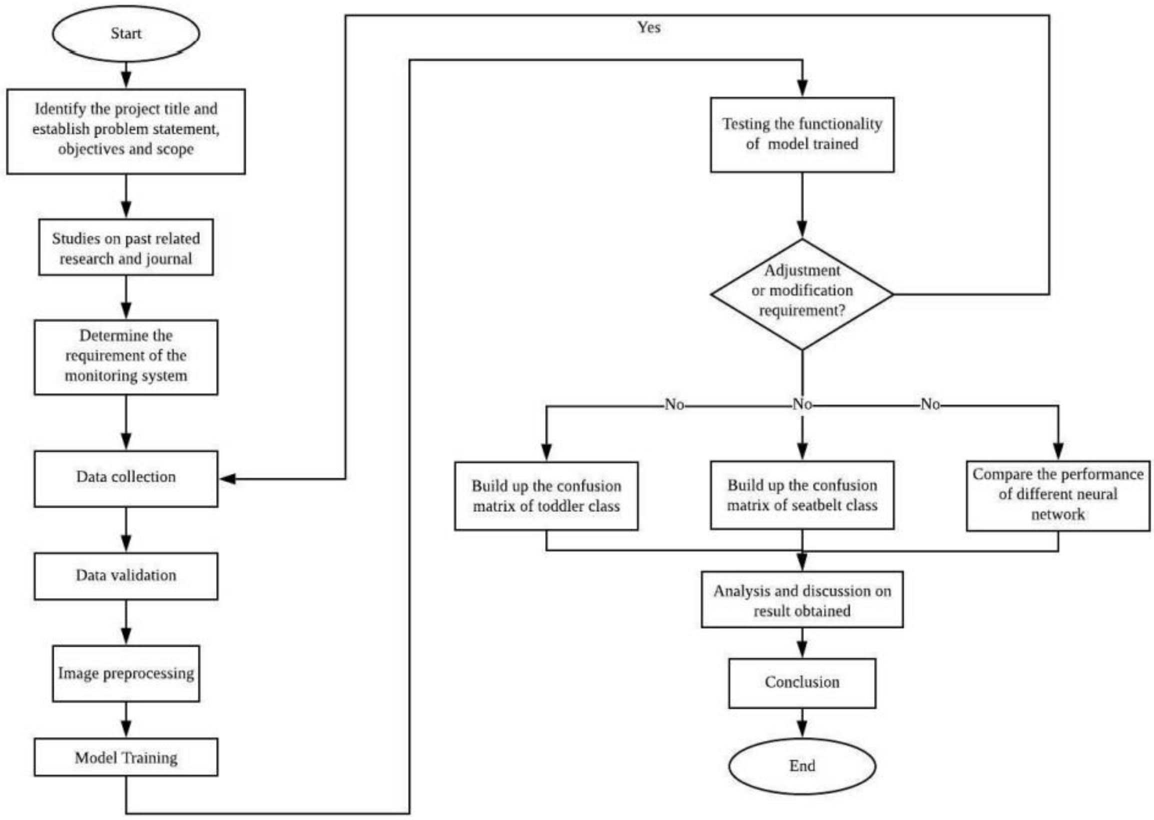


Figure 2. Flowchart of the toddler monitoring system

3.1. Hardware selection

Due to the rapid growth of technologies nowadays, many types of single-board computers such as Raspberry Pi, Intel and Nvidia can be found on the market. The Nvidia Jetson Nano was chosen because of the application programming interface (API) model created by Nvidia, which is compute unified device architecture (CUDA). CUDA is a parallel computing platform and API model. It enables developers to use the CUDA-enabled graphic processing unit (GPU) for general-purpose processing, allowing the term general-purpose computing on graphics processing units (GPU) to its full extent. Developers can significantly accelerate computing applications by leveraging the power of GPUs and the presence of CUDA.

3.2. Algorithm development

The flow of the design of the development for the monitoring system is shown as in Figure 3. Firstly, the image of the toddler in the backseat will be collected and the image will be preprocessed before labelling to ensure all the images are the same in type and size. After the annotation is done, the images will be fed into the system for training. The trained system will test for functionality.

Figure 3. Flowchart of toddler monitoring system design

3.3. Data collection and annotation

The data is collected in video form (.mp4) and all the videos are extracted into images for training purposes. The location to collect the data is set at the back of the seat. Examples of the collecte images are shown as in Figure 4.

Figure 4. Sample images collected

Data annotation is the process of adding metadata to a dataset in preparation for training a machine learning model. This process is to generate an annotation file that contains the information about the box location of the region and the name of the annotation for all the images. The function of the annotation file is to help machines learn certain patterns and correlate the results. LabelImg is used as a graphical image annotation tool as shown as in Firgure 5. It can output an annotation file in a Pascal VOC XML file. The annotation makes two classes called "toddler" and "seatbelt" that can find out if a "toddler" or "seatbelt" is present.

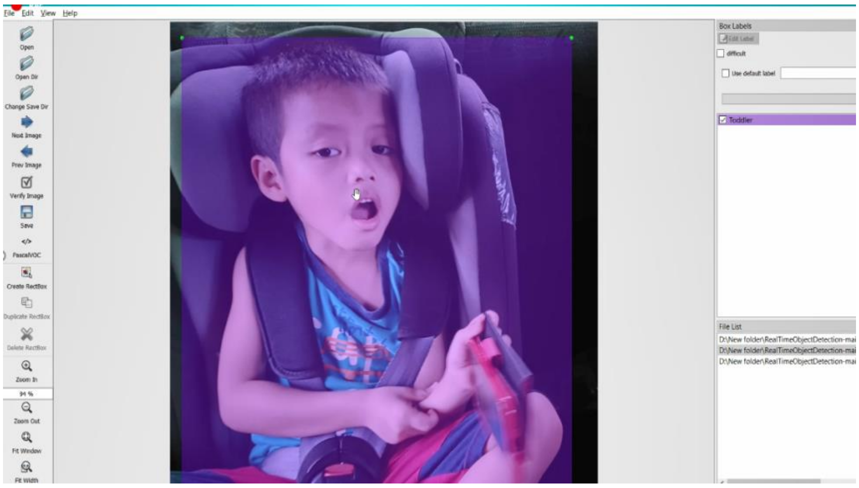


Figure 5. Labelimg annotation process

3.4. System training

The network was trained on Google Colab. The provided GPU was used to train the model. The model is initialized with the original SSD-Mobilenet and SSD-Inception. Only the output layers were pre-trained. The sample image of the output from the trained network is shown in Table 1. The situation is classified by checking the number of bounding boxes from different classes on the image as shown in Table 2.

Table 1. Sample of image result from trained network

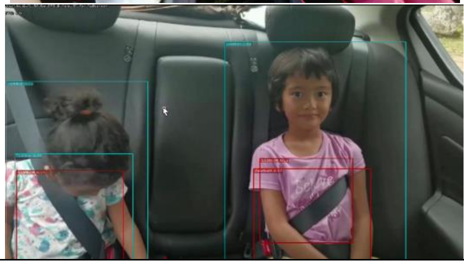
SSD-Mobilenet	**SSD-Inception**
	
	

Table 2. Sample of image result for classified images

Backseat condition/output of system	Image Result
Toddler is absent.	
One or more than one toddler is not using seatbelt. (Speaker starts to beep to alert driver)	
All the detected toddlers are using the seatbelts.	

4. RESULT

The results are compared by using the confusion matrix method. Table 3 shows the confusion matrix table for performance comparison. The parameters such as the Accuracy, Precison and Recall are calculated based on the (1)-(3). Tables 3(a) and 3(b) are the result of the performance that is calculated by using the data from the confusion matrix table and the summary in graph format as shown as in Figure 6.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Table 3. Confusion Matrix of (a) SSD-Mobilenet and (b) SSD-Inception

SSD-MOBILENET									
TODDLER					SEATBELT				
		PREDICTED		TOTAL			PREDICTED		TOTAL
		NO	YES				NO	YES	
ACTUAL	NO	3	3	6	ACTUAL	NO	50	1	51
	YES	41	141	182		YES	31	106	137
TOTAL		44	144	188	TOTAL		81	106	188
CORRECTLY DETECTED			75%		CORRECTLY DETECTED			56.38%	

(a)

SSD-INCEPTION									
TODDLER					SEATBELT				
		PREDICTED		TOTAL			PREDICTED		TOTAL
		NO	YES				NO	YES	
ACTUAL	NO	5	1	6	ACTUAL	NO	50	1	51
	YES	41	141	182		YES	35	102	137
TOTAL		46	142	188	TOTAL		85	103	188
CORRECTLY DETECTED			75%		CORRECTLY DETECTED			54.26%	

(b)

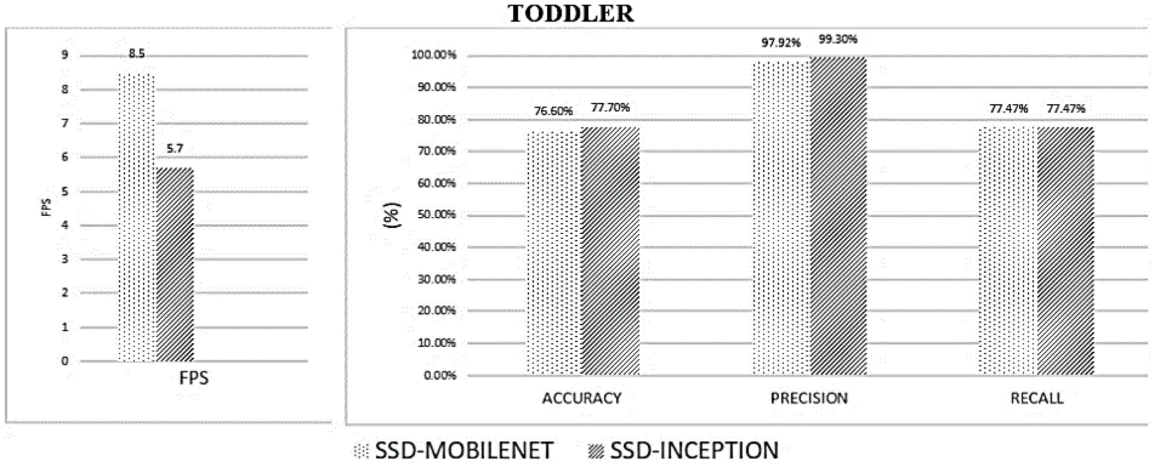
From Table 4, it was found that SSD-Inception gives better performance with 77.70% of accuracy, 97.92% of precision and 77.47% of recall when detecting the toddler class, while SSD-Mobilenet performs better in the class of seatbelt with 82.98% of accuracy, 99.07% of precision and 77.37% of recall. Though there is a performance difference between both neural networks, it is just a slight difference. SSD-Mobilenet has a higher frame per second (FPS) which is 8.5 FPS, than SSD-Inception, which has 5.7 FPS. It means that SSD-Mobilenet can respond faster than SSD-Inception as the performance of both neural networks is only slightly different in accuracy. The performance comparison between the networks is shown in Table 5 and later by Figure 6 (a) & (b). The tensorboard function is used to get the mean average precision (mAP) with 0.5 intersection over union (IoU) of both neural networks as shown in Table 5.

Table 4. Performance of neural network

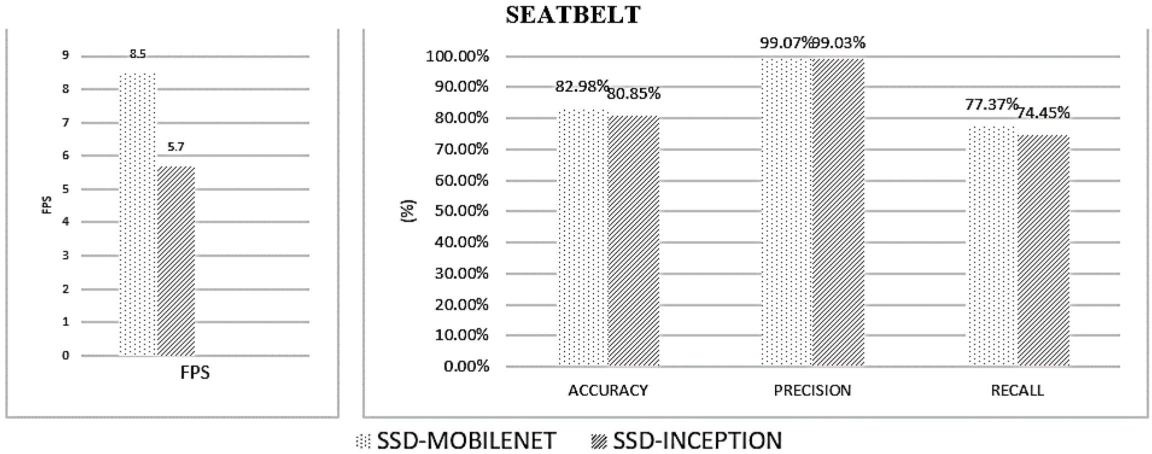
TYPE OF NEURAL NETWORK	TODDLER FPS	TODDLER ACCURACY	TODDLER PRECISION	TODDLER RECALL	SEATBELT FPS	SEATBELT ACCURACY	SEATBELT PRECISION	SEATBELT RECALL
SSD-MOBILENET	8.5	76.60%	97.92%	77.47%	8.5	82.98%	99.07%	77.37%
SSD-INCEPTION	5.7	77.70%	99.30%	77.47%	5.7	80.85%	99.03%	74.45%

Table 5. Performance of neural network

Average Precision	Classes		Mean Average Precision
	Seatbelt	Toddler	
SSD-Mobilenet	0.94129	0.980041	0.973779
SSD-Inception	0.88132	0.927863	0.904591



(a)



(b)

Figure 6. The performance of neural network (a) Toddler class and (b) Seatbelt class

5. CONCLUSION

This paper presents comprehensive work on the design and development of a toddler monitoring system to determine the seatbelt condition of the toddler in the backseat and inform the driver about the safety condition of the toddler. The toddler monitoring system is designed to be vision-based and the neural network method is used. The Jetson Nano is used as a microcontroller for the system due to its powerful performance to run the neural network for object detection. The SSD-type neural network is the best choice for Jetson Nano because it needs less processing power from the mobile controller. In the SSD-type neural network, SSD-Inception and SSD-Mobilenet are chosen and compared. The comparison of the performances of different neural networks has been carried out, and the result is shown in the previous chapter. It can be concluded that SSD-Mobilenet has better performance in speed, which is FPS when processing a video image, while the accuracy of both neural networks has no large difference. As the work progressed at this stage, several future expansion and development ideas were noted. For future improvements, the system's accuracy and sensitivity need to be improved by using more different data with a different model to the vehicle, toddlers with different ages, skin color, and so on to increase the database. Furthermore, the system can interact with the vehicle to get a more accurate output. To achieve this, cooperation with vehicle companies needs to be conducted.

ACKNOWLEDGEMENTS

The authors would like to express greatest appreciation to Universiti Teknologi Malaysia (UTM) and Ministry of Higher Education (MOHE), Malaysia for the financial support (grant number: Q.K130000.2456.08G28). We also acknowledged the financial from UTEM in this research work conducting this research.

REFERENCE

[1] O. C. Ann and L. B. Theng, "Human activity recognition: A review," *Proceedings - 4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014*, pp. 389–393, 2014, doi: 10.1109/ICCSCE.2014.7072750.

[2] H. B. Zhang *et al.*, "A comprehensive survey of vision-based human action recognition methods," *Sensors (Switzerland)*, vol. 19, no. 5, 2019, doi: 10.3390/s19051005.

[3] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, "Vision-based human activity recognition: A survey," *Multimedia Tools and Applications*, vol. 79, no. 41–42, pp. 30509–30555, 2020, doi: 10.1007/s11042-020-09004-3.

[4] S. Ranasinghe, F. Al MacHot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation," *International Journal of Distributed Sensor Networks*, vol. 12, no. 8, 2016, doi: 10.1177/1550147716665520.

[5] J. Ondruš, E. Kolla, P. Vertaľ, and Ž. Šarić, "How do autonomous cars work?," *Transportation Research Procedia*, vol. 44, pp. 226–233, 2020, doi: 10.1016/j.trpro.2020.02.049.

[6] P. A. Hancock, I. Nourbakhsh, and J. Stewart, "On the future of transportation in an era of automated and autonomous vehicles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 16, pp. 7684–7691, 2019, doi: 10.1073/pnas.1805770115.

[7] J. Deng, X. Xuan, W. Wang, Z. Li, H. Yao, and Z. Wang, "A review of research on object detection based on deep learning," *Journal of Physics: Conference Series*, vol. 1684, no. 1, 2020, doi: 10.1088/1742-6596/1684/1/012028.

[8] I. E. N. and M. J. Murphy, "What is machine learning?," *Springer International Publishing*, pp. 3–11, 2015, doi: 10.1007/978-1-4842-3799-1_3.

[9] J. Bell, "What is machine learning?," in *Machine Learning and the City*, Wiley, 2022, pp. 207–216.

[10] D. T. Pham and P. T. N. Pham, "Artificial intelligence in engineering," *International Journal of Machine Tools and Manufacture*, vol. 39, no. 6, pp. 937–949, 1999, doi: 10.1016/S0890-6955(98)00076-5.

[11] P. Gambus and S. L. Shafer, "Artificial intelligence for everyone," *Anesthesiology*, vol. 128, no. 3, pp. 431–433, 2018, doi: 10.1097/ALN.0000000000001984.

[12] A. A. Goshtasby, "CEG 724-01 : Computer vision I CEG-724 computer vision I," *Computer science and engineering syllabi*, pp. 1–3, 2007.

[13] D. Forsyth and J. Ponce, "Computer vision: A modern approach. (Second edition) - archive ouverte HAL," *Prentice Hall*, vol. 17, pp. 21–28, 2003, [Online]. Available: https://hal.archives-ouvertes.fr/hal-01063327/.

[14] R. Szeliski, "Computer vision: Algorithms and applications," *Springer Science & Business Media*, 2010.

[15] I. Mihajlovic, "Everything you ever wanted to know about computer vision," *Analytical and Bioanalytical Chemistry*, vol. 391, no. 7, pp. 2373–2376, 2008.

[16] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[17] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 747–750, 2020, doi: 10.1007/s13042-020-01096-5.

[18] R. U. S. A. Data, F. Application, P. Data, D. Dynamic, S. Detectors, and I. Appli, "United States patent (19)," no. 19, 1998.

[19] C. C. Tappert, "Who is the father of deep learning?," *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019*, pp. 343–348, 2019, doi: 10.1109/CSCI49370.2019.00067.

[20] GeeksforGeeks, "Difference between artificial intelligence and automation," 2022.

[21] S. Yan, Y. Teng, J. S. Smith, and B. Zhang, "Driver behavior recognition based on deep convolutional neural networks," *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016*, pp. 636–641, 2016, doi: 10.1109/FSKD.2016.7603248.

[22] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B. Y. Choi, and T. R. Faughnan, "Real-time human detection as an edge service enabled by a lightweight CNN," *Proceedings - 2018 IEEE International Conference on Edge Computing, EDGE 2018 - Part of the 2018 IEEE World Congress on Services*, pp. 125–129, 2018, doi: 10.1109/EDGE.2018.00025.

[23] H. Q. Bao, H. Q. Bao, and C. Sun-tae, "A light-weight gender / age estimation model based on multi-taking deep learning for an embedded system," pp. 1–4, 2020.

[24] P. Murthy, O. Kovalenko, A. Elhayek, C. Gava, and D. Stricker, "3D human pose tracking inside car using single RGB spherical camera," *ACM Chapters Computer Science in Cars Symposium (CSCS-17)*, 2017, [Online]. Available: http://av.dfki.de.

[25] J. J. Yan, H. H. Kuo, Y. F. Lin, and T. L. Liao, "Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing," *Proceedings - 2016 IEEE International Symposium on Computer, Consumer and Control, IS3C 2016*, pp. 243–246, 2016, doi: 10.1109/IS3C.2016.72.

[26] E. Jose, M. Greeshma, T. P. Mithun Haridas, and M. H. Supriya, "Face recognition-based surveillance system using FaceNet and MTCNN on Jetson TX2," *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, pp. 608–613, 2019, doi: 10.1109/ICACCS.2019.8728466.

BIOGRAPHIES OF AUTHORS

Kok Jia Quan received the bachelor's degree in Mechatronics Engineering in 2021 from Universiti Teknikal Malaysia Melaka (UTeM). Experience as an Automatic Testing Machine (ATE) development Engineer at Aimflex Sdn.Bhd. Currently, work as SCADA Engineer at A-Control & Instrumentation Pte.Ltd and can be contacted at jiaquan97919@gmail.com.

Zamani Md Sani received his degree in 2000 from Universiti Sains Malaysia. He worked at Intel Malaysia Kulim for 6 years and obtained his Master at the same university later in 2009. Later he joined education at Universiti Teknikal Malaysia Melaka and obtained his PhD from Multimedia Universiti in 2020. His research interest is in Image Processing and Artificial Intelligence and can be contacted at zamanisani@utem.edu.my.

Tarmizi Bin Ahmad Izzuddin received his doctoral degree from Universiti Teknologi Malaysia (UTM), and currently serving as the head of the Rehabilitation Engineering and Assistive Technology (REAT) research group at Universiti Teknikal Malaysia Melaka (UTeM) His research interest includes Neural Network Algorithms, Brain-Computer Interfaces and Robotics, particularly Neurorobotics. He holds multiple professional AI Engineering certificates, including certification from IBM and can be contacted at tarmizi@utem.edu.my.

Azizul Azizan obtained his B.Eng. (Hons.) Electronics Engineering (Telecommunications) degree from Multimedia University. He received his PhD qualification in 2009, from University of Surrey in the area of 3.5G physical layer adaptation for satellite systems. He is currently with the Advanced Informatics Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia and can be contacted at azizulazizan@ieee.org.

Hadhrami Ab Ghani received his bachelor degree in electronics engineering from Multimedia University Malaysia (MMU) in 2002. In 2004, he completed his master's degree in Telecommunication Engineering at The University of Melbourne. He then pursued his Ph.D. at Imperial College London in intelligent network systems and completed his Ph.D. in 2011. He can be contacted at email: hadhrami.ag@umk.edu.my. His current research interests are advanced communications, network security and computer vision. Currently he is a senior lecturer at Faculty of Data Science and Computing, Universiti Malaysia Kelantan. He can be contacted at email: hadhrami.abdghani@mmu.edu.my.